DOCUMENT RESUME

ED 148 738 SP 011 634

AUTHOR McDonald, Frederick J.
TITLE The State of the Art in Performance Assessment of Teaching Competence.
PUB DATE Feb 73
NOTE 19p.; Presented to the Multi-State Consortium, American Educational Research Convention (New Orleans, Louisiana, February 1973)

EDRS PRICE MF-$0.83 HC-$1.67 Plus Postage.
DESCRIPTORS *Ability Identification; Educational Assessment; Educational Problems; Effective Teaching; *Evaluation Criteria; *Program Development; *Research Needs; *Teacher Evaluation; *Teaching Skills

ABSTRACT

The author argues that, while substantial problems exist in the comprehensiveness, cost, and complexity of an evaluation system for performance assessment, the effort should be undertaken and problems attacked and solved as they arise. Three questions are investigated in describing the current state of performance assessment: (1) What are the goals of systems measuring teaching competence? (2) What progress has been made in developing procedures required for such systems? and (3) What problems remain to be solved in creating effective, efficient systems for competence assessment? Six goals are identified: validity, reliability, "learnability" of skills, freedom from bias, comparability and standardization, and sensitivity to opportunities to learn skills. These goals also represent the procedural broblems to be overcome. One suggested method for solving them is through the use of simulated teaching conditions. Through the use of this technique, skills likely to be useful in teaching would be identified. Student teaching and internship experiences would also be useful to assess daily performance under uncontrolled conditions, as would simulation devices. Theoretical problems are more complex and require extensive research and development. The most practical way of working at present is for state education department personnel to set standards for the evidence about competence that is to be provided, however the competence may be defined, rather than think about assessment in terms of what competencies should be measured. (MJB)

****
* Documents acquired by ERIC include many informal unpublished *
* materials not available from other sources. ERIC makes every effort *
* to obtain the best copy available. Nevertheless, items of marginal *
* reproducibility are often encountered and this affects the quality *
* of the microfiche and hardcopy reproductions ERIC makes available *
* via the ERIC Document Reproduction Service (EDRS). EDRS is not *
* responsible for the quality of the original document. Reproductions *
* supplied by EDRS are the best that can be made from the original. *
****

PB0958

ED148738

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY
Educational Testing Service
TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC) AND THE ERIC SYSTEM CONTRACTORS."

U.S. DEPARTMENT OF HEALTH, EDUCATION & WELFARE
NATIONAL INSTITUTE OF EDUCATION
THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

THE STATE OF THE ART IN PERFORMANCE ASSESSMENT OF TEACHING COMPETENCE

Frederick J. McDonald

Educational Testing Service
Princeton, New Jersey

Performance-Based Teacher Education Project
AACTE

Paper presented to the Multi-state Consortium at the American Educational Research Association Convention
New Orleans
February 1973

SP011 634

To describe and evaluate the state of the art in performance assessment of teaching competence, we should answer three questions: (1) What are the goals of the systems to measure teaching competence? (2) What progress have we made in developing the procedures required for systems measuring teaching competence? (3) What problems must be solved to create effective and efficient systems for assessing teaching competence?

Before attempting to answer these questions, I should first openly acknowledge the assumptions that influence my ideas and those that do not. I assume that even small progress made in assessing teacher competence will be a great improvement over our present evaluations. Because I assume this, I am willing to urge the use of procedures and systems which at the present time are limited or even defective--since I also assume that as we use what instruments and techniques are now available we shall learn more about the nature of teaching competence and progressively improve our methods for evaluating it. I further assume that the cost of measuring teacher competence will not necessarily be prohibitive, despite the fact that many people seem to assume the contrary. It is obvious that instituting a measurement system where there was none before will impose heavy costs. But until we have a complete cost/benefit analysis, I am not willing to assume that these costs will be unmanageable. Further, I do not assume that the evaluation system will necessarily be complex because we do not know as yet what information is needed to make reliable and valid judgments about teaching competence. I choose, therefore, to be optimistic about both the cost and complexity of an evaluation system.

It seems to me that the most reasonable position to take at this time is to acknowledge the many problems and to regard their solution as dependent

PB0958

upon the ordinary processes of research and development. It is self-defeating and destructive to so negatively prejudge the situation that work which might resolve the problems is either impeded or prematurely terminated.

The Goals of an Evaluation System

Let us now return to the three questions with which I began this presentation. The first is obviously the most important: what should be the goals of a performance assessment system? The ultimate goal of such a system is to provide information that enables the decision makers--teacher educators, certification officers, and administrators--to decide whether a teacher candidate has sufficient competence to be permitted to teach. Everything else that we will want to say about this goal is either a refinement of what is meant by making a judgment about competence, or a problem that must be solved if such judgments are to be made. I will therefore list a set of conditions that must be satisfied for this goal to be attained.

First, measurement procedures used in the evaluation of teaching competence must have high validity. Validity in this case means we have demonstrated a substantial relationship between a teaching skill or performance and its effects upon students.

At this point in time we know very little about what skills or performances have demonstrable effects on student behavior. The highest priority in a research program on teaching competence must be given to solving this problem. The conclusion, however, should not be drawn that we must defer the development of an evaluation system until all the relevant research has been done. We already have an abundance of ideas on pertinent

teaching competencies which we can begin to measure--a necessary first step, and one whose effect on teaching performance can be studied systematically as part of the process of developing evaluation systems. We cannot or ought not, however, make definitive judgments about teaching competence until the validities of specific teaching skills or performances have been established.

Second, we must establish the reliability of assessment procedures that will be used in evaluation systems. Reliability in this case means the stability of a teaching skill or performance. Does the teaching skill vary considerably from day to day, class to class, topic to topic? We need to know the conditions under which it is most likely to be variable and those under which it is most likely to be constant.

While considerable attention is given to the problem of validity, practically no attention has been given to the performance characteristics of teaching skills. There is some intuitive understanding that a variety of conditions may affect a teacher's skill, but formal research on the problem is nonexistent. It makes no difference how research on this aspect of teaching performance comes out because we need all kinds of information in making judgments about competence. If, for example, a particular teaching skill requires special conditions to elicit it, we must ask whether those conditions were present when the teacher's possession of the skill was being evaluated.

Third, we need to know what, for lack of a better term, I will call the "learnability" of a skill. It is conceivable that a particular skill, even though it has been demonstrated to have highly desirable effects upon children's learning, can be learned by only a few individuals. We

do not expect everyone to become theoretical physicists or concert pianists or gold medal-winning athletes because we know that some special aptitudes may be required for achievements of this order. Similarly, there may be forms of teaching whose acquisition demands unusual aptitudes. Thus, we should avoid the trap of assuming that every skill demonstrated to have desirable effects on students will necessarily be required of every person who wishes to teach.

We must develop the kinds of information described in these three points--relating to the reliability, validity, and learnability of teaching skills. We cannot achieve our goals unless we can define what is to be measured, under what conditions it is to be measured, and how the information yielded by the procedure is to be interpreted. But there are some other conditions beyond these three that must be met by any system for measuring teaching competence.

A fourth condition, for example, is that the information gathered about teachers must be as uncontaminated by subjective biases and political processes as is humanly possible. Any observation of human behavior is, of course, influenced to a degree by the characteristics of the observer and by the characteristics of the measurement procedure itself. We shall need to develop new strategies for collecting information on teaching competence if we are to minimize the problems of gathering such information.

Fifth, the conditions of measurement must provide comparable information on groups of teachers. We must standardize the conditions under which teacher behavior is measured. Otherwise, it is impossible to tell whether variations in teacher behavior are due to situational differences or to differences in teacher competence, thus exposing the candidate to the vagaries of idiosyncratic favoritism.

The conclusion should not be drawn, however, that the assessment system should ignore variations in the conditions under which teacher performance occurs. As I stated earlier, it is essential that we determine those conditions under which a teaching skill may be expected to vary. Such information is necessary in evaluating teaching competence. But we cannot treat teaching as if it were so different on each separate occasion that we can never evaluate it. The conflict between establishing reasonable expectations for teaching performance and the variety and complexity of the situations in which teaching occurs is one of the most important problems we have to solve. Until it is solved, our decisions about competence must necessarily be tentative.

Sixth, we must develop an evaluation system that makes due allowance for the teachers' opportunities to acquire the relevant teaching skills. A decade may well pass before teacher education programs have become performance-based and before a reasonable amount of performance assessment has taken place. To demand competence of individuals who have not had the opportunities to acquire that competence is immoral, and to make judgments that affect their careers on such an insubstantial basis is equally immoral. Assessment should <u>not</u> be used for evaluative purposes when the individuals being assessed have not had the opportunity to learn those things for which they are being evaluated.

In listing these constraints, I have outlined the major problems that must be solved if a system of assessing teaching competence is to be developed and used in a practical way; indeed the solution of the problems depends on our meeting these conditions.

The State of Measurement Technology and Knowledge

Despite popular opinion to the contrary, measurement technology follows the wants of its users. For three-quarters of a century, decision-makers of one kind or another have wanted to assess what candidates for teaching positions know. Measurement technology for assessing academic knowledge thus became highly developed. Consequently, we now have widely available tests of knowledge of subject matter and of knowledge about teaching methods. But until recent years there has been no demand of any consequence for the development of performance assessment. As a consequence, this aspect of measurement is largely undeveloped.

We have developed observation scales which permit us to report what teachers do under ordinary teaching conditions. But none of these observation schedules or methods meet any of the constraints which I outlined previously. The conclusion should not, however, be drawn that the use of observation schedules should be terminated; nevertheless, they and any other measurement procedure must be subjected to the kinds of research and development necessary to make them useful within the stipulated constraints.

The basic problem of performance assessment technology is to create a controlled situation or to use an ordinary teaching situation with sufficient knowledge of its characteristics so that their influence on the performance can be assessed. Observation schedules fall into this second group of procedures. But the problem of assessing the influences in a wide variety of teaching conditions and performances is enormous. I have come to the conclusion that a more profitable line of research should use controls, simulated teaching situations.

Rather than describe the adequacy of available measurement technology,

a rather depressing picture, I shall describe the work in progress on the assessment of performance using simulated teaching conditions. This work is relatively new in the field of teaching competence, and it is conceivable that it may never develop into a practicable technology. But some of the problems must be solved and are more likely to be solved efficiently and satisfactorily by using simulated teaching conditions. The relationships between teaching performance and student learning, for example, can be studied very efficiently by using microteaching sessions. Similarly, the reliability of teaching performances under controlled conditions can also be easily assessed.

We are attempting to build a system that provides information about basic teaching skills, teaching performances, and teaching strategies. This system is designed to use simulated teaching situations of varying degrees of complexity. Initially, both the teaching skill measured and the conditions under which it is measured are relatively simple. The prospective teacher moves through the system from this simple initial level to one where he or she must use a variety of skills, performances, and strategies in combination and must enact them over a period of time. The critical feature of the component systems is that the conditions are carefully controlled. In what ways are they controlled? For any given component, the teaching objectives and teaching performances are specified; for example, if we are attempting to assess the teacher's ability to reinforce students participating in a discussion, the objective of the lesson is set as increasing the amount of student participation in the class. The trainee is told that he will be evaluated on the degree to which students participate in a class discussion. Initially, we want to

measure only his use of reinforcement procedures to facilitate participation in classroom discussion. In other components and at other times the teaching situation will be constructed to evaluate the trainee's ability to present a stimulating problem for discussion or to ask questions which elicit comments from students or facilitate discussion by highlighting differences of opinion.

In order to make assessments at this time certain factors have to be controlled. For example, if we are to study the teacher's use of reinforcement techniques, we do not want the teacher preoccupied with the preparation or development of the topic or with the choice of the topic, nor do we want a group of teachers to have different objectives. Therefore, as I noted, we stipulate the objective, and we also provide the teaching materials for the lesson. Further, we control variability of the students by reassigning the students to a teaching assignment and by providing the preparation period for familiarizing themselves with the material to be discussed. It should be noted, and I wish to emphasize, that the factors which are controlled in one type of situation may be systematically varied in another in order to assess the degree of the teacher's skill.

What can be learned by using this technology? These microteaching sessions can be used to assess a teacher's skill under a variety of different teaching conditions. We have tried several formats of microteaching lessons to determine what can be larned by having teachers teach different lessons and different groups of students. In one format, for example, we asked the teacher to teach four different topics to the same group of students; in another, a teacher teaches different students the same topic; and in yet another, the teacher will teach different topics

to the same students and then the same topics to another set of students. In these ways we can test how the teacher's skills change under the different teaching conditions. We have already found that inexperienced teachers' behavior varies from topic to topic, to some extent, even though the topics appear to be highly similar in nature and would seem to offer similar opportunities for using teaching skills.

Eventually, we would expect to have a series of representative teaching situations in which it had been shown that certain kinds of teaching skills are highly likely to be useful. An assessment battery would consist of a series of microteaching sessions covering these teaching situations.

Because the microteaching format has several limitations, such as relatively short lessons and small numbers of students used, we developed a minicourse format to use for more complex teaching situations. A minicourse in this context is a short course lasting one or two weeks. The teachers teach this course an hour a day for one of those periods of time. For these courses we have also developed a complete statement of objectives, course materials, and achievement tests to be taken by the students at the end of the course. Almost three hundred preservice teachers have taught these courses so that we have gathered considerable data on teaching performance and experience with the technology itself.

The results of the analyses of these teaching performances indicate that the microteaching performances are relatively poor predictors of the teaching performances in the minicourses. But, for several reasons, I am not willing to offer this conclusion as solidly substantiated. First, the subject-matter of the short courses is more complex, and segments of any one lesson would be similar to the teaching situations created by the

microteaching session. But their combination presents a different kind of teaching problem. Also, in the situation in which we have gathered data, relatively little performance training took place between the initial microteaching and the subsequent teaching in the short courses. Systematic training using the assessment gathered in the microteaching might influence substantially the relationships between the two kinds of performances; that is, where the teacher has sufficient skills, as revealed in the microteaching performance, the instructors would reinforce these skills so that they were more likely to be used by trainees in the more complex teaching. There is still reason to believe that we have created two sufficiently distinctive types of performance situations which will yield equally valuable but different kinds of information. My own opinion is that the microteaching is most useful for assessing the degree to which a teacher has basic skills; the minicourse is most useful in assessing how teachers integrate these skills into complex teaching performances.

As I said earlier, our goal is to build a system that will assess the simple components of teaching competence and also its more complex aspects. We visualize the coordinated assessment system as made up of periods of microteaching used primarily for assessment purposes and short courses or units that can later be embedded in internship and student-teaching experiences. These components would be arranged so that the training institution could be given information continuously about the level of competence being reached by the trainee.

The student teaching and internship experience can be used to assess daily performance under uncontrolled conditions. They are useful for providing information on what teachers are likely to do in contrast to

what they are able to do. The simulation-type teaching situations that I have described should be thought of as teachers assessing what teachers are capable of doing; whereas on-the-job performance tells us what he or she regularly does. We can also use on-the-job observation to assess such factors as teaching style.

I am convinced, after a year's experience with administering systems such as these and another year's analysis of the data gathered, that it is entirely feasible to build a useful performance assessment system that can be embedded in any teacher-training program. As such systems are developed, state departments of education can learn what criteria ought to be applied to the assessment of teaching competence. The prototype that we are developing in our research program can be used to generate concepts about the appropriate means to evaluate each training institution's procedures for assessing the competence of their students.

It is obvious to me that the type of techniques being developed can be used with beginning teachers prior to the granting of tenure, and for periodic reassessment of teaching skill. We should not, however, use these procedures at the present time for continuing certification schemes, any more than we ought to use the current observational procedures and administrator evaluations for that purpose. But there is good reason to believe that, in time, assessment strategies can be developed which will be useful both in preservice and inservice assessment of teaching competence.

One question asked frequently is, "How expensive will performance assessment be?" As I observed earlier, there is no adequate way now to answer this question because we do not know what information we shall need to make judgments about teaching competence. But my experience indicates

that we can gather large amounts of data about teaching performance relatively inexpensively; for example, three twenty-minute microteaching sessions yield an hour of teaching performances. A videotape of these performances can be observed to assess many different skills. In fact, the short course yields almost more information than can be quickly processed. We have resorted to audiotaping and videotaping about a half hour of each hour session for the very practical reason that this half hour covers three-fourths of the teaching that occurs. But I am not convinced that even this much information is needed. We are studying this problem by taking different samples of the videotape performances that we have and finding the intercorrelations among teaching performances across the samples. This technical work will eventually help us find the optimum amount of observation time that we must use in order to make reliable judgments about teaching skills.

While many people are concerned about the cost of performance testing, few discuss the economies that may be achieved by using them. It is very apparent that, when one looks at videotapes of teaching performances, a diagnosis of basic teaching skills can be made very easily, and this information can be used to speed the learning of these teaching skills. We already have sufficient information on feedback and modeling techniques to know that the information generated in a performance assessment system can be incorporated into feedback and modeling systems to increase the effectiveness of training. Thus, we may expect the institution to benefit by achieving economies in training time and the trainee to benefit by achieving competence sooner when he enters upon actual teaching.

Reactive Simulation Devices

I have been describing performance assessment procedures which require the teacher to interact with students in a simulation of actual teaching. There are some other procedures that appear to be promising in assessing potential teaching competence. We have developed a filmed simulation test that portrays a teacher conducting a class. The film is stopped periodically and the viewer is asked to say what he or she would do in this situation; in other parts of the test the viewer is asked to explain what is occurring in the class, and, in some places, he is asked what advice or suggestions he would give to the teacher. We do not yet have data on the predictive validity of this measurement, but it obviously taps the prospective teacher's attitudes towards students, his perceptions of the problems of teaching, and his understanding of some phases of teaching strategies. The test situation is sufficiently lifelike to create great interest on the part of those who view the film, and invariably the film produces some intense discussions about what the teaching portrays, thus suggesting to us that the film also has great potential for training purposes.

Another type of simulation device has been created to depict the kinds of strategies teachers use. This procedure is a gamelike device in which an experimenter plays the role of student in a highly controlled way. The teacher arranges the subject matter in the form of presentations or questions, and the experimenter responds whenever the teacher asks a question. This gamelike situation does discriminate sharply between deductive and inductive teaching styles. It seems likely that the role of the experimenter can be computerized so that the game can be used by large numbers of trainees in a more efficient manner.

Both of these techniques are illustrative of the kinds that are likely to be developed to assess teacher competence. Obviously, their validity in predicting teaching performance has to be established. But the work to date suggests that they are probably of sufficient validity to be useful in assessing what a teacher is likely to do in a classroom.

The Problems to be Solved

The work that I have described here is meant to be illustrative of the possibilities of assessing teaching competence. I emphasized that the problems are primarily technical and partially theoretical. We have not yet developed an adequate technology of performance assessment, but the reason for this is related more to the lack of experience with these technologies than to a lack of knowledge of what to do.

The theoretical problems are more complex and will require extensive research and development. The kinds of procedures that I have been talking about here can be adapted to experiments--for example, assessing the relationship of teaching skills to pupil learning. Carefully controlled experiments can be designed in the microteaching format to study hypotheses about these relationships. The minicourse data can be used to induce hypotheses about the relationship of teaching competence to pupil learning. But until the basic questions about what competencies make a difference have been resolved, the design of assessment technology will be incomplete.

Each of the conditions governing an assessment system, that I outlined previously, presents a technical problem to be solved. These problems are highly amenable to research, and there is no reason to believe that they cannot be solved in a relatively short period of time.

A basic problem is political in character. Many individuals believe that competence cannot be measured because they do not want competence to be measured. The solution to this problem is obviously complex. Certainly the development of efficient and demonstrably useful assessment systems will weaken the arguments against assessment. Still, there will probably always be those who will say that the most important aspects of teaching are unmeasurable. But I, for one, have never been able to understand how this conclusion can be reached before anybody has tried seriously to measure the more elusive aspects of teaching.

Conclusion

What should state departments of education, who wish to move to performance-based certification, do about the assessment problem? Obviously, an enormous amount of research and development work remains to be done, and even though we are optimistic about the pace at which the research and development can proceed, we cannot at this time offer state department personnel finished performance assessment systems. For that matter, we all know that just selecting the competencies that will be the criterion performances cannot be done at the present time.

I have the impression that state department personnel are largely concerned, and justifiably so, about this latter question. I propose that rather than think about assessment in terms of what competencies should be measured, state department personnel think about the criteria for evaluating assessment systems from which judgments about competence are made. If you agree with me that it will be several years before we have a good body of research established on the relationship of teaching competence to pupil learning, then it seems to me the most practical course to take is to set

standards for the evidence about competence that is to be provided, however that competence may be defined.

I urge your consideration of this approach because I believe that once a faculty in an institution becomes concerned with the assessment problem, the definition of competence and the demonstration of its relevance to student learning will become major concerns of the faculty and will more likely be tackled quickly and efficiently. But it also seems to me that what state department personnel ought to be asking is what the bases are to be on which judgments about competence are made. If attention is focused on these decision-making processes and the information is used to make decisions about competence, we shall at least have guaranteed that the judgments are not being made subjectively or that a multiplicity of factors affecting competence have been inadequately controlled or that the range of a teacher's skill has not been appropriately evaluated. The problem for the next several years can be stated as a question addressed to the training institution, "Given your conception of what constitutes competence, what evidence have you gathered that demonstrates that teachers have acquired these competencies?"

In my opinion, if state departments would adopt criteria for the evidence to be presented to support teacher certification, the performance-based movement would move forward quickly. You would also avoid the trap of falling back into approving programs which have a performance base but cannot produce evidence that the trainees who go through the program are in fact competent teachers. This approach would focus the institution's attention on the development of the assessment of competence and would give each institution greater freedom in the development of programs.

Those of us who are interested in the performance-based movement in teacher education accept a basic principle: we do not care how a teacher becomes competent provided that he or she can demonstrate that competence. If we believe in this principle, as I know we do, then our efforts should be directed to deciding what evidence we will assess to determine whether a teacher is competent.